Dynamic Probing for Intrusion Detection under Resource Constraints

Keqin Liu*, Qing Zhao*, Ananthram Swami†

*Department of Electrical and Computer Engineering
University of California, Davis, CA 95616, {kqliu,qzhao}@ucdavis.edu

†Army Research Laboratory, Adelphi, MD 20783
Email: ananthram.swami@us.army.mil

***Abstract*—We consider a large-scale cyber network with N components. Each component is either in a healthy state or an abnormal state. To model scenarios where attacks to the network may not follow a stochastic process and the attackers may adapt to the actions of the intrusion detection system (IDS) in an arbitrary and unknown way, we adopt a non-stochastic model in which the attack process at each component can be any unknown deterministic sequence. Due to resource constraints, the IDS can only choose K $(K < N)$ components to probe at each time. An abnormal component incurs a cost per unit time (depending on the criticality of the component) until it is probed and fixed. The objective is a dynamic probing strategy under the performance measure of regret, defined as the performance loss compared to that of a genie who knows the entire attack processes *a priori* and probes optimally (under certain constraints) based on this knowledge. We propose a policy that achieves sublinear regret order, thus offers the same time averaged performance as that of the omniscient genie.**

***Index Terms*—Intrusion detection, dynamic probing, non-stochastic multi-armed bandit, regret.**

I. Introduction

Persistent monitoring of systems connected to the internet (both wired, and now increasingly wireless and hybrid) is critical to the maintenance of security and privacy in the face of threats from increasingly sophisticated adversaries. Intrusion detection systems (IDS) have long been accepted as an essential layer in providing security to the information infrastructure.

With the increasing size, diversity, and interconnectivity of the cyber system, however, intrusion detection faces the challenge of scalability: how to accurately and rapidly detect and locate intrusions and anomalies in a large dynamic network with limited resources. The two basic approaches to intrusion detection, namely, active probing and passive monitoring [1], [2], face stringent resource constraints when the network is large and dynamic. Specifically, active-probing based approaches need to choose judiciously which paths and components of the network to probe to reduce overhead (see, e.g., [3]); passive-monitoring based approaches need to determine how to sample the network so that real-time processing of the resulting data is within the computational capacity of the IDS [4]. The problem is compounded by the fact that models of adversarial behaviors are typically unknown, non-parametric, and evolving.

[0]This work was supported by Army Research Lab under Grant W911NF1120086.

A. Resource-Constrained Dynamic Probing under Unknown Non-Stochastic Attacks

In this paper, we consider intrusion detection in a large dynamic cyber network where, due to resource constraints, the IDS cannot monitor (through either active probing or passive sampling) all the components in the network simultaneously. At each time instant, the IDS can only monitor part of the network (similar models have also been considered in [5]–[7]). The objective is to choose judiciously which components of the network to probe to reduce the overall cost incurred to the network by all infected components. Our focus is on handling adversarial behaviors that are unknown, non-stationary, and potentially reactive to the strategies of the IDS.

Consider a network with N components which can be paths, routers, or subnets. At a given time, a component can be in a healthy state or an abnormal state. An abnormal component remains abnormal until the anomaly is detected and resolved. A healthy component may be attacked and become abnormal if the attack is successful. To model scenarios where attacks to the network may not follow a well-behaved stochastic process and the attackers may adapt to the actions of the IDS in an arbitrary and unknown way, we adopt a non-stochastic model in which the attack process at each component can be any unknown deterministic sequence. Equivalently, we aim to minimize the network cost for any possible attack traces. The results in this paper thus apply to arbitrary adversary models. It also applies to the scenario where the IDS only sees one trace (or a small number of traces) of attack realizations, and policies designed for ensemble-average performance under a stochastic model may not be applicable.

The above problem can be considered as a variation of the non-stochastic multi-armed bandit (MAB) problem first considered by Auer *et al.* [8]. In an MAB problem, there are N arms and a single player. Under the non-stochastic modek, the reward offered by each arm when played is given by an arbitrary unknown deterministic process. At each time, the player chooses K $(K < N)$ arms to play and accrues the current rewards from the chosen arms. The performance

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Dynamic Probing for Intrusion Detection under Resource Constraints	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Davis,Department of Electrical and Computer Engineering,Davis,CA,95616**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
to appear in Proc. of IEEE International Conference on Communications (ICC), June, 2013

14. ABSTRACT

We consider a large-scale cyber network with N components. Each component is either in a healthy state or an abnormal state. To model scenarios where attacks to the network may not follow a stochastic process and the attackers may adapt to the actions of the intrusion detection system (IDS) in an arbitrary and unknown way, we adopt a non-stochastic model in which the attack process at each component can be any unknown deterministic sequence. Due to resource constraints, the IDS can only choose K (K < N) components to probe at each time. An abnormal component incurs a cost per unit time (depending on the criticality of the component) until it is probed and xed. The objective is a dynamic probing strategy under the performance measure of regret, de ned as the performance loss compared to that of a genie who knows the entire attack processes a priori and probes optimally (under certain constraints) based on this knowledge. We propose a policy that achieves sublinear regret order, thus offers the same time averaged performance as that of the omniscient genie.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**5**	

measurer of an arm selection policy is given by *regret*, defined as the performance loss compared to an omniscient genie who knows the entire reward process of every arm *a priori* and plays optimally based on this knowledge. To minimize regret, or equivalently, to approach the performance of the omniscient genie, it is crucial that the player learns from past reward observations and improves its arm selection over time. However, when the reward process of each arm is an arbitrary deterministic sequence, past reward observations bears no information on the current reward, and the regret will grow linearly when the genie can always choose the largest reward among all arms at each time. To make the performance measure of regret meaningful, Auer *et al.* imposed constraints on the genie so that approaching the average performance of the genie becomes possible. This leads to the so-called weak regret, in which, the genie that the player is competing with, while still knows the entire reward process of every arm noncausally, can only choose one fixed arm to play over the entire time horizon. In this case, what the player is trying to learn is which arm has the largest *cumulative* reward rather than trying to catch the largest reward at each time instant. Intuitively, the former is possible as past reward observations become increasingly more informative for learning the largest cumulative reward as time goes. Indeed, as shown in [8], arm selection policies can be constructed to achieve a weak regret of order $O(\sqrt{T})$, which is sublinear with the time horizon length T, indicating that the time-average performance of the omniscient genie under switching constraints can be approached as T increases. Auer *et al.* further considered a more general switching constraint (referred to as the hardness constraint) that specifies the maximum allowable number of arm switchings that the genie can take over the entire horizon T. They constructed a learning policy that achieves regret $O(\sqrt{TS(T)})$ under a hardness constraint $S(T)$. The regret is thus sublinear when the genie can only switch at a sublinear rate with the horizon length T.

The intrusion detection problem considered in this paper can be formulated as a variation of the non-stochastic MAB where each component is an arm associated with an unknown cost process. However, there are two major differences between our problem and the non-stochastic MAB studied in *et al.* [8]. First, in the non-stochastic MAB, rewards are accrued only over the chosen arms, while in the intrusion detection problem, costs occur over all components that are abnormal even when they are not chosen. Second and more importantly, the cost process of each component is determined by both the unknown attack process and the actions of the IDS, whereas in the non-stochastic MAB, the reward processes are independent of the player's actions. As a consequence, the current action of the IDS not only affects the immediate performance loss compared to the genie, but also affects future losses. To see this, consider a case where at time t, the genie probed and fixed an abnormal component while the IDS probed a normal one. The impact of this action cannot be clearly bounded since the unfixed abnormal component will continue incurring cost until it is probed. These two major differences make a direct extension of [8] inapplicable, and indeed, we show in this paper that the regret performance in the intrusion detection problem can be drastically different from that given in [8].

Our main results are as follows. First, by constructing a specific policy, we show that the weak regret of the intrusion detection problem can grow at an arbitrarily slow rate in T (i.e., arbitrarily close to a *bounded* weak regret or, in other words, complete learning). This is in sharp contrast with the $O(\sqrt{T})$ weak regret order that was shown to be optimal for non-stochastic MAB in [8]. We further show that any bounded weak regret cannot be achieved. The proposed policy is thus order optimal. Under a general hardness constraint $S(T)$ on the genie, we show that the proposed policy achieves regret $O(T^{2/3}S^{1/3}(T))$, which is sublinear when $S(T)$ is sublinear. When all the components incur the same cost when infected, the regret reduces to $O(\sqrt{TS(T)})$.

B. Related Work

Existing work on dynamic probing for intrusion detection often assumes a stochastic model for the attack processes. For example, in [9], a Markovian attack model was adopted and the intrusion detection problem was addressed based on Markovian decision processes and Q-learning. In [10], a more general, potentially non-Markovian stochastic attack model was considered. There are also quite a few studies on statistical modeling of the attack processes [11]–[14] that aims to extract the statistic model of the attacks based on the observed data of IDS.

II. Problem Statement

Consider a network with N components. The attack process at a component n is given by an arbitrary unknown deterministic sequence denoted by $\{a_n(t)\}_{t\geq 1}$:

$$a_n(t) = \begin{cases} 1, & \text{if attack is successful;} \\ 0, & \text{otherwise.} \end{cases}$$

Each component has two sates, healthy (0) or abnormal (1). Under a successful attack, the component state $x_n(t)$ becomes abnormal and stays abnormal until it is probed and fixed (see Fig. 1 for an illustration). An abnormal component n incurs a cost c_n per unit time. The objective of the IDS is to minimize the total network cost over a horizon T by judiciously choosing K $(K < N)$ components to probe at each time.

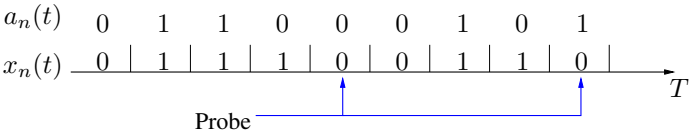

Fig. 1. The attack process $a_n(t)$ and the state $x_n(t)$ of a component n.

Let $u_n(t) = \{1 \text{ (probe)}, 0 \text{ (not probe)}\}$ denote the action applied to component n at time t. Let $C_n(t)$ denote the cost incurred by component n at time t. The process $\{C_n(t)\}_{t\geq 1}$ is determined by both the unknown attack process $\{a_n(t)\}_{t\geq 1}$ and the actions $\{u_n(t)\}_{t\geq 1}$ of the IDS on component n.

Specifically, $C_n(t) = c_n$ if there exists a $t_0 \leq t$ such that $a_n(t_0) = 1$ and $u(\tau) = 0$ for $\tau = t_0, t_0+1, \ldots, t$, and $C_n(t) = 0$ otherwise (see Fig. 1).

We measure the performance of a probing policy π by regret, defined as the total performance loss compared to the optimal policy π_S^g of a genie who knows non-causally the entire attack process on every component but has a hardness constraint $S(T)$. Let $R_S^\pi(T)$ denote the regret of policy π up to time T. We have

$$R_S^\pi(T) = \mathbb{E}_\pi \left[\sum_{t=1}^{T} \sum_{n=1}^{N} C_n^\pi(t) - \sum_{t=1}^{T} \sum_{n=1}^{N} C_n^{\pi_S^g}(t) \right],$$

where $\mathbb{E}_\pi[\cdot]$ denotes the expectation under policy π which may be a randomized policy, and $C_n^\pi(t)$ the cost incurred by component n at time t under a policy π.

It is easy to see that regret is nondecreasing with time. The objective is to minimize the growth rate of regret $R_S^\pi(T)$ with respect to time T under any possible attack processes $\{a_n(t)\}_{t\geq 1, 1\leq n\leq N}$. Note that regret is a finer performance measure compared to the time-average cost criterion since any sublinear regret order leads to the minimum average cost achieved by the genie as time goes to infinity. A smaller regret order leads to a faster convergence to the minimum average cost.

In Sec. III, we show that if $S(T)$ is linear with T, then a sublinear regret is not achievable; otherwise we propose a policy to achieve a sublinear regret.

III. Sublinear Regret under Hardness Constraint

In this section, we consider the case where the regret is defined with respect to a genie that can switch at a rate of $S(T)$ (i.e., a hardness constraint of $S(T)$). The case of weak regret where the genie cannot switch is addressed in Sec. IV.

Under a hardness constraint $S(T)$ on the genie, we first establish a lower bound on regret, which implies that if $S(T)$ has a linear order with T, then the corresponding regret under any policy π grows at least at a linear order with time. We then propose a policy, referred to as Clear-Stay-Explore (CSE), that achieves a sublinear regret as long as the hardness constraint $S(T)$ is sublinear with T.

A. A Lower Bound on Regret

Theorem 1: The regret order is lower bounded by the order of the hardness constraint $S(T)$, *i.e.,* we have[1]

$$R_S^\pi(T) = \Omega(S(T)).$$

Proof: Construct an attack process as follows. First, the number of attacks up to time T is set equal to $S(T)$. Second, at each time at most K components can be attacked. Third, at each time, given that M $(M \leq K)$ components will be attacked, these M components are drawn uniformly from the set of N components.

Based on the above attack process, we consider the expected total cost under any policy π where the expectation is taken over all realizations of attacks. Clearly, at each time when at least one attack occurs, the expected immediate cost under any policy π is lower bounded by a constant due to the uniform drawing of the attacked components. Consequently, the total expected cost of policy π grows at least at the same order of $S(T)$ since the number of attacks is equal to $S(T)$. On the other hand, the genie achieves zero total cost since at most K components can be attacked at each time and the total attacks are equal to the hardness constraint $S(T)$. The expected regret of policy π thus grows at least at the same order of $S(T)$, where the expectation is again taken over all realizations of attacks. Therefore, there must exist one realization of attacks where the regret grows at least at the same order of $S(T)$. ■

B. The CSE Policy

Since sublinear regret is not achievable when $S(T)$ is linear, we will focus on the case where $S(T)$ is sublinear with time. In the following, we proposed the CSE policy to achieve a sublinear regret when $S(T)$ is sublinear. For the ease of the presentation, we present the policy for the case of $K = 1$ (i.e., can only probe one component at each time). The extension to the general case is straightforward.

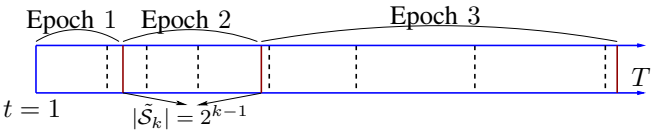

Fig. 2. The epoch structure of the CSE policy (dotted line: virtual switching times in $\tilde{\mathcal{S}}(T)$, $|\tilde{\mathcal{S}}_k|$ denotes the number of virtual switching times in epoch k).

In order to achieve a sublinear regret order and approach the genie's time-average performance over time, it is crucial that the policy be able to localize and follow (to a certain quantifiable degree) the switching actions of the genie which depend on the attack processes unknown to the policy. This is possible due to the sublinear hardness constraint on the genie. The basic idea of the proposed CSE policy is thus to first construct a sequence of virtual switching times for the purpose of localizing the switching actions of the genie. Specifically, let $\tilde{\mathcal{S}}(T)$ be a sequence of time instants from 1 to T, which we refer to as the virtual switching times. This sequence can be arbitrary as long as its cardinality grows with T at the same order[2] as $S(T)$:

$$|\tilde{\mathcal{S}}(T)| = \Theta(S(T)).$$

The basic structure of the CSE policy is constructed based on this sequence of virtual switching times $\tilde{\mathcal{S}}(T)$. Specifically, we first partition time into epochs, where the kth epoch contains 2^{k-1} virtual switching times in $\tilde{\mathcal{S}}(T)$ (see Fig. 2). This construction of the epoch structure ensures that the ratio of the true switching times of the genie to the virtual switching times

[1]For any two functions $f(T)$ and $g(T)$, $f(T) = \Omega(S(T))$ is equivalent to $\limsup_{T\to\infty} \frac{f(T)}{S(T)}$ being positive, possibly infinite.

[2]For any two functions $f(T)$ and $g(T)$, $f(T) = \Theta(g(T))$ is equivalent to $\limsup_{T\to\infty} \frac{f(T)}{S(T)}$ being positive and finite.

is uniformly bounded by a constant in all epoches and under any attack processes (see Lemma 1). This allows us to design the actions of the IDS that leads to bounded performance loss compared to the genie within each epoch.

We then further partition each epoch into segments with equal length (see Fig. 3). The length of the segments, which depends on the epoch length, will be optimized for minimum regret as given in Theorem 2. Within each segment, the actions of the policy can be summarized as clear, stay, and explore. Specifically, at the beginning of each segment, we probe all components in a round-robin fashion to **clear** all abnormal components caused by previous attacks. After clearing all the component, we continue the round-robin search and **stay** at the first component that becomes abnormal. In the remaining of the segment, we **explore** periodically (see Fig. 3) where the exploration period depends on the segment length and will be optimized for minimum regret (see Theorem 2). In each exploration, we check only those components that have a higher cost (more critical) than the previous component we are staying with. Such components are probed one by one starting from the most critical one for at most one round. We stop at the first such component that is found to be abnormal and stay with this component (that ends the current exploration). If no such components are found to be abnormal during this one round of exploration, we go back to the component we were previously staying with until the next exploration. This completes the description of the CSE policy.

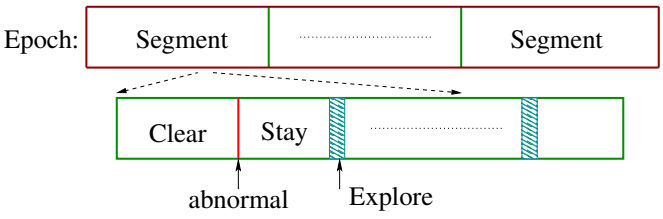

Fig. 3. The structure of each epoch.

C. The Regret Analysis

In this section, we analyze the regret performance of the CSE policy and specify the optimal choice of the parameters in the CSE policy.

Theorem 2: Let T_k denote the length of the kth epoch. The optimal length $T_s^*(T_k)$ of segments in this epoch is given by

$$T_s^*(T_k) = \lfloor (\frac{T_k}{|\tilde{\mathcal{S}}_k|})^{2/3} \rfloor,$$

where $|\tilde{\mathcal{S}}_k|$ denotes the number of virtual switching times in epoch k. The optimal exploration period T_x^* in each segment is given by

$$T_x^*(T_k) = \sqrt{T_s^*(T_k)}.$$

The resulting CSE policy has the following regret:

$$R_S^\pi(T) = O(T^{2/3} S^{1/3}(T)).$$

The optimal value of the segment length $^*_s(T_k)$ is to ensure that within each epoch, there is only a sublinear number of segments that contain the genie's switchings. In other words, as time goes, in most of the segments in each epoch, the genie does not switch and probes the same component. Regret in such segments is lower, making a sublinear regret order possible. The optimal exploration period T_x^* is obtained by balancing the tradeoff between the delay in catching an attack to a more critical component and the overhead associate with exploring.

The proof is based on the following two lemmas. The first one shows that for all epochs, the number of genie switchings is within a constant factor of the virtual switching times, *i.e.,* the sequence of virtual switching times $\tilde{\mathcal{S}}(T)$ gives an order-accurate estimate of the number of genie's switchings in each epoch. This is important in analyzing the regret in each epoch, as given in the second lemma. The detailed proof is omitted.

Lemma 1: Let A_k denote the number of genie's switchings in the kth epoch, which depends on the underlying attack processes unknown to the policy. Let $|\tilde{\mathcal{S}}_k|$ denotes the number of virtual switching times in epoch k. We have, for all k,

$$\frac{A_k}{|\tilde{\mathcal{S}}_k|} \leq C$$

for some constant C independent of k and the underlying attack processes.

Lemma 2: Let T_k denote the length of the kth epoch. The regret in this epoch is given by

$$R_S^\pi(T_k) = O(T_k^{2/3} (A_k)^{1/3}),$$

where A_k is the number of genie's switchings in this epoch.

We now consider a special case where all components have the same cost c_n. In this case, we show that the regret can be improved with two minor modifications to the CSE policy. First, we partition the kth epoch into $\sqrt{T_k\,|\tilde{\mathcal{S}}_k|}$ segments. Second, we do not need to carry out exploration in each segment.

Theorem 3: If all components have the same cost $c_n \equiv c$, the modified CSE policy achieves regret

$$R_S^\pi(T) = O(\sqrt{TS(T)}).$$

IV. Achieving Order-Optimal Weak Regret

In this section, we consider weak regret, *i.e.,* the genie cannot switch and always stays on the same set of K components. We show that a slight modification of the CSE policy achieves a regret arbitrarily close to finite. We then establish a strict lower bound on regret showing that a finite regret cannot be achieved under any policy. The modified CSE policy is thus order-optimal.

In the modified CSE policy, we do not partition time into epochs or segments. We first carrier out a round-robin search and stay with the first K components that are found to be abnormal. We then perform the same exploration procedure with a rate of $E(T)$ (i.e., total $E(T)$ explorations over a horizon of length T) that grows at an arbitrarily slow rate with T.

Theorem 4: Under the modified CSE policy, the regret has order $E(T)$, which can be set arbitrarily close to $O(1)$. Furthermore, a finite regret cannot be achieved under any policy.

V. Conclusion

We studied the intrusion detection problem under an unknown non-stochastic attack model. We proposed the CSE policy to achieve a sublinear regret order with respect to an omniscient genie who knows noncausally the entire attack processes.

References

[1] S. Jajodia, P. Liu, V. Swarup, and C. Wang, *Cyber Situational Awareness*, Springer, 2009.

[2] H. Debar, M. Dacier, and A. Wespi, "Towards A Taxonomy of Intrusion-Detection Systems," *Computer Networks*, vol. 31, no. 8, pp. 805-822, 1999.

[3] P. Barford, N. Duffield, A. Ron, and J. Sommers, "Network Performance Anomaly Detection and Localization," *Proc. of INFOCOM*, June, 2009.

[4] M. Kodialam and T. V. Lakshman, "Detecting Network Intrusions via Sampling: a Game-Theoretic Approach," *Proc. of INFOCOM*, 2003.

[5] I. Kuwatly, M. Sraj, Z. Al Masri, and H. Artail, "A Dynamic Honeypot Design for Intrusion Detection," *Proc. IEEE Int'l. Conf. Pervasive Services (ICPS)*, 2004.

[6] G. Vigna and R.A. Kemmerer, "NetSTAT: A Network-based Intrusion Detection Approach," *Proc. 14th IEEE Computer Security Applications Conference*, 1998.

[7] P. Brutch, and C. Ko, "Challenges in intrusion detection for wireless adhoc networks", *Proc. Symp. Applications and the Internet Workshops*, 2003.

[8] P. Auer, N. Cesa-Bianchi, Y. Freund, R. E. Schapire, "The Nonstochastic Multiarmed Bandit Problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48-77, 2003.

[9] T. Alpcan and T. Basar, "An Intrusion Detection Game with Limited Observations," *Proc. Of 12th International Symposium on Dynamic Games and Applications*, 2006.

[10] K. Liu and Q. Zhao, "Dynamic Intrusion Detection in Resource-Constrained Cyber Networks," in *Proc. of IEEE International Symposium on Information Theory*, July, 2012.

[11] D.E. Denning, "An Intrusion-Detection Model," *IEEE Transactions on Software Engineering*, no. 2, vol. SE-13, pp. 222-232, 1987.

[12] M. Roesch, "Snort-Light Weight Intrusion Detection for Networks," *Proceedings of the 13th Large Installation System Administration Conference*, 1999.

[13] A. K. Ghosh, A. Schwartzbard, and M. Schats, "Learning Program Behavior Profiles for Intrusion Detection," *Proceedings of the 1st conference on Workshop on Intrusion Detection and Network Monitoring*, 1999.

[14] T. Bass, "Intrusion Detection Systems and Multisensor Data Fusion," *Communications of ACM*, no. 4, vol. 43, April, 2000.